一种新的提高互联网端到端时延精度的测量方法*)

潘 乔 裴昌幸 朱畅华
(西安电子科技大学综合业务网 ISN 国家重点实验室 西安 710071)

摘 要 由于端到端网络时延的测量中存在收发时钟不同步的问题,在测量中大多是通过测往返时延来间接求得端到端时延,测试结果误差较大。本文利用主动探测方法,在互联网上通过在一端发送带有时间戳的 IP 数据包,在另一端记录该测量分组的到达时间戳来获得端到端的时延测量值,然后利用线性规划的方法来消除了收发时钟的初始相位差和相对频差等影响,计算出网络的端到端时延真实值。通过实例测试,结果表明该方法消除了时钟不同步带来的误差,提高了测试结果的精确度。

关键词 端到端时延,优化算法,时钟同步,主动测试

A Novel Method for Improving the Precision in Internet End-to-End Delay Measurements

PAN Qiao PEI Chang-Xing ZHU Chang-Hua
(National Key Lab. of Integrated Service Networks, Xidian Univ., Xi'an 710071)

Abstract The clock of the sending endpoint is always nonsynchronous with that of the receiving one in Internet end-to-end delay measurement, so the delay is usually obtained indirectly by measuring the round-trip delay time, but the result is quite inaccurate. We propose a novel method to measure the Internet end-to-end delay directly. First this paper describes an active measurement method to get the time difference between the time a sending endpoint sends an IP packet and the time a receiving endpoint receives the IP packet. And then a linear programming method is given that to remove the skew and offset between two host clocks. The results show that the method works well.

Keywords End-to-end delay, Linear programming, Clock synchronization, Active measurement

1 引言

端到端时延是评估 Internet 的网络性能的重要参数之一。通过对端到端时延的测量,能够分析当前 Internet 的基本特性,如网络的拓扑结构和网络的流量模型等,为网络技术的改进提供可靠的理论依据。例如,在 VoIP[1]应用中需要提供端到端时延的限制(如当单向时延超过 250ms,VoIP 应用的性能将大打折扣);了解网络中时延的分布,可根据 Internet 的 QoS 要求,对网络进行规划设计。

常用的网络时延测量方法只能测出往返时延,例如用 Ping 命令来测网络的时延,它只能测网络的往返时延。由于网络链路的非对称性,不能简单地用除以 2 的方法来获得端到端时延,加之其性能和 TCP、UDP 或其他 IP 协议有一定的出入(路由器给 ICMP 协议的优先性较低),因此测得的数据有一定的局限性,时间精度只能在毫秒级。如果利用直接发送数据包的方法,在互联网上通过从一端发送带有时间戳的 IP 数据包,在另一端记录该测量分组的到达时间戳来获得端到端的时延,由于存在收发时钟不同步的问题,它只是一个测量值。

目前大多数测量均借助于 GPS 接收机或 NTP(Network Time Protocol)协议来实现对于收发主机时钟的同步[2]。但是,采用 GPS 接收机不但价格贵,且与接收环境有关,NTP 精度又不够(毫秒级)。本文利用线性规划的方法消除了收发时钟的初始相位差和相对频差,计算出网络的端到端时延真实值来实现端到端时延的测量。它不依赖于 GPS 接收机且由于我们直接提取主机的时钟,时间精度可以达到微秒级。

2 时延的组成

定义(端到端时延) 在端到端网络时延测量配置中,发送固定大小的 IP 探测数据包,源端离开的时标和到达目的端的时标之差称为 IP 探测数据包的端到端时延[3]。

概括地说,网络端到端时延被分为 4 个主要部分:处理时延、传输时延、传播时延和排队时延。

令最小要求的数据包的大小为 M_0 比特(仅是头字节加上尾字节的大小),总的数据包的大小为 M 比特。在第 i 个节点,R_i 为传输速度,则分别用 $T_{p,i}$, $T_{t,i}$, $T_{ew,i}$, $T_{q,i}$ 表示处理时延大小、传输时延大小、从节点 i 到节点 j 的传播时延大小、M 比特大小的数据包的排队时延大小。R 为总的传输速度,L 为传播时延和处理时延的总和,T_q 为排队时延的总和,则对于一个 k 跳的链路:

$$L=\sum_{i=1}^{k}T_{ew,i}+\sum_{i=1}^{k-1}T_{p,i},R=\sum_{i=1}^{k}R_i,T_{t,i}=\frac{M}{R_i},T_q=\sum_{i=1}^{k}T_{q,i}$$

M 比特大小的数据包的端到端时延:

$$delay_t(M)=(L+\frac{M_0}{R})+(T_q+\frac{M-M_0}{R})$$

式子中第一项$(L+\frac{M_0}{R})$是端到端时延的固定部分,它依赖于最初设计的网络性能(如数据包的传输速度)和最短的数据包长 M_0;第二项$(T_q+\frac{M-M_0}{R})$是端到端时延的随机部分,它依赖于网络的拥塞程度(如链接利用率)和数据包大小 M。

*)国家自然科学基金重点项目(60572147,60132030)。潘 乔 博士研究生。

3 常用的端到端时延的测量方法的局限性

由于用 Ping 命令来测出网络的往返时延，然后计算出端到端时延的方法存在着缺陷，因此在端到端时延测量中主要是直接测量网络的端到端时延[4]。但是在端到端时延测量配置中，测试点往往位于不同的地点，如果直接测量网络端到端时延，存在收发主机时钟不同步的局限性。在 IP 数据包中的是各自主机的时钟时间，两个主机时钟的频率和初始相位不同，如果没有时间同步机制，会造成测量数据的错误。我们发送数据包来测试源主机(西电科大校园网)到目的主机(中国电信数据网)的端到端时延。测试结果见图 1，横轴为发送分组序列号，纵轴为测得的时延(单位为微秒 μs)，收发时钟分别采用各自的时钟，由于在源主机和目的主机之间时钟不同步(存在频差)的影响，测试的结果有一个线性的趋势。

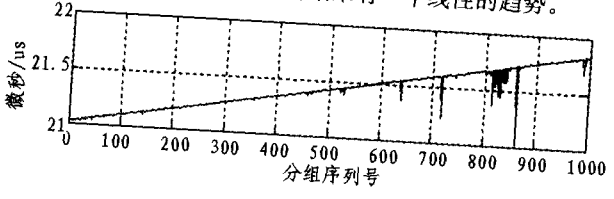

图 1　端到端时延实测结果

4 时延测量方法的分析及高精度算法的实现

4.1 端到端网络时延的测量方法

目前有许多工具可以测试端到端网络的时延[5]，总体上分为两类：

一类是主动探测方法，即测试端到端网络时延通过传输测试数据包，但存在收发时钟不同步的问题；另一类是被动探测方法，即从网络上截取数据包进行分析，用被动探测方法来分析端到端的时延，它需要在每个终端机上执行服务终端程序，这对于普通的网络用户是不可能实现的[6]。

本文采用主动探测方法，在互联网上通过在一端发送带有时间戳的 IP 探测数据包，在另一端记录该测量分组的到达时间戳来获得端到端的时延测量值，计算出网络链路的前向、反向和往返时延值(未同步)。发送端主机 A 发送带有时间戳的 IP 探测数据包，记录下发送时标 $T_{i,0}$，而接收端主机 B 在接收到 IP 探测数据包后记录下接收时标 $T_{i,1}$；另一方面，接收端主机 B 同时回复一个带有时间戳的 IP 探测数据包，并记录回复时标 $T_{i,2}$，原发送端主机 A 接收到该回复包记下接收时标 $T_{i,3}$；由此可得到在网络上的 3 个时延时间(未同步)：

(1)前向时延：$T_{i,1}-T_{i,0}$；

(2)反向时延：$T_{i,3}-T_{i,2}$；

(3)往返时延：$T_{i,3}-T_{i,0}$；

如图 2，其中① $T_{ew,i}$，② $T_{t,i}$，③ $T_{p,i}$，④ $T_{q,i}$。

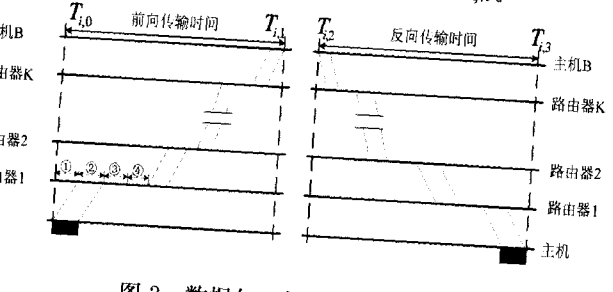

图 2　数据包 i 在网络中的传输

4.2 时延测量中时钟不同步影响的消除

Sue B. Moon 提出的时钟同步的方法[7]是目前比较常用的方法，它消除了源主机和目的主机之间时钟的频率差，但是没有解决收发时钟的频率差和初始相位差引起的时间偏移的问题。本文通过对端到端网络时延测量的数据分析，同时对反向链路也进行测量，就可以消除收发时钟的频率差和初始相位差，解决了时间偏移的问题，提高了端到端时延测量的精度。

设发送端主机时钟的初始相位为 t_{0s}(此时真实时间为 s_0)，频率为 $\lambda_s, s_1, s_2, \cdots, s_n$ 为发端时钟以 t_{0s} 为参考点、在计数时对应于标准时钟的相应时刻值，发送时的标准时间为 t_s，故发送主机的时钟函数 $T_s(t)$ 为

$$T_s(t_s)=t_{0s}+\int_{s_0}^{s_1}\lambda_s\,dt+\int_{s_1}^{s_2}\lambda_s\,dt+\cdots+\int_{s_n}^{t_s}\lambda_s\,dt$$

设接收端主机时钟的初始相位为 t_{0d}(此时真实时间为 d_0)，频率为 $\lambda_d, d_1, d_2, \cdots, d_n$ 为收端时钟以 t_{0d} 为参考点、在计数时对应于标准时钟的相应时刻值，接收的标准时间为 t_d，故接收主机的时钟函数 $T_d(t)$ 为

$$T_d(t_d)=t_{0d}+\int_{d_0}^{d_1}\lambda_d\,dt+\int_{d_1}^{d_2}\lambda_d\,dt+\cdots+\int_{d_m}^{t_d}\lambda_d\,dt$$

此时时延的真实值为

$$delay_t=t_d-t_s$$

而测量得到的值(未同步)：

$$\begin{aligned}delay_m&=T_d(t_d)-T_s(t_s)\\&=\left(t_{od}+\int_{d_0}^{d_1}\lambda_d\,dt+\int_{d_1}^{d_2}\lambda_d\,dt+\cdots+\int_{d_m}^{t_d}\lambda_d\,dt\right)-\\&\quad\left(t_{os}+\int_{s_0}^{s_1}\lambda s\,dt+\int_{s_1}^{s_2}\lambda_s\,dt+\cdots+\int_{s_n}^{t_s}\lambda_s\,dt\right)\end{aligned}$$

整理得

$$\begin{aligned}delay_t=&\frac{1}{\lambda_d}delay_m-\frac{\lambda_d-\lambda_s}{\lambda_d\lambda_s}(T_s-t_{0s})\\&-\left[\frac{t_{0d}-t_{0s}}{\lambda_d}+(s_0-d_0)\right]\end{aligned}\qquad(1)$$

上式中，为书写方便，用 T_s 代替 $T_s(t_s)$。

如果以接收端的时钟为参考，则 $\lambda_d=1$，若 $d_0=t_{0d}=0$，此时

$$delay_t=delay_m-\frac{1-\lambda_s}{\lambda_s}(T_s-t_{0s})+t_{0s}-s_0$$

其中 $t_{0s}-s_0$ 即为收发时钟的偏差。上式可写为

$$delay_t=delay_m=-\beta\cdot T_{sn}+\varphi$$

其中 $T_{sn}=T_s-t_{0s}$；　$\beta=\dfrac{1-\lambda_s}{\lambda_s}$；　$\varphi=t_{0s}-s_0$

测量反向时延，由(1)式可直接写为

$$delay_n=\frac{1}{\lambda_s}delay'_m-\frac{\lambda_s-\lambda_d}{\lambda_s\lambda_d}(T'_d-t_{0d})-\left[\frac{t_{0s}-t_{0d}}{\lambda_s}+(d_0-s_0)\right]$$

其中带“′”的量为对应的反向链路的值。

如果以接收端的时钟为参考时钟，则 $\lambda_d=1$，若 $d_0=t_{0d}=0$，此时

$$delay_n=\frac{1}{\lambda_s}delay'_m+\beta T'_d-\beta t_{0s}-\varphi$$

利用前向和反向两个单向测量可以计算出消除收发时钟频差和初始偏差的端到端的时延。对于前向链路：

优化目标为

$$\min\sum_{n=1}^{N}(delay_{mn}-\tilde{\beta}T_{sn}-\hat{\xi}\cdot M_n+\hat{\theta})$$

约束条件为

$$delay_{mn}-\tilde{\beta}T_{sn}-\hat{\xi}\cdot M_n+\hat{\theta}\geqslant 0,\hat{\beta},\hat{\xi},\hat{\theta}\geqslant 0$$

其中：$\hat{\beta},\hat{\xi},\hat{\theta}$ 为估计值。$delay_m$ 为实测到的发送第 n 个数据包的时延值。

$T_{sn}=T_s-t_{os}$ （T_s 为发送主机的时钟函数，t_{os} 为发送主机时钟的初始相位）

$\hat{\beta}=\frac{1-\lambda_s}{\lambda_s}$ （λ_s 为发送端主机的时钟频率）

$\hat{\xi}=\sum_{n=0}^{N}\frac{M_n}{b_n}$ （M_n 数据包的大小，b_n 带宽）

$\hat{\theta}=\varphi-T_c$ （$\varphi=t_{os}-s_0$，s_0 为起始 $T_c=T_{ew}-T_p$）

对于反向链路：

优化目标为

$\min\sum_{n=1}^{N}\left(\frac{1}{\hat{\lambda}_s}delay'_{rm}-(\hat{\lambda}_s-1)T'_d-\hat{\xi}'\cdot M_n-\hat{\eta}\right)$，其中 $\hat{\eta}=\frac{1-\hat{\lambda}_s}{\hat{\lambda}_s}t_{0s}+\hat{\varphi}+\hat{T}'_c$。

约束条件为

$\frac{1}{\hat{\lambda}_s}delay'_{rm}-(\hat{\lambda}_s-1)T'_d-\eta-\hat{\xi}'\cdot M_n\geqslant 0$。假设被测链路往返方向的传播和处理时延和相等，则 $\hat{T}'_c=\hat{T}_c$。设发端初始相位为 0（如在计算机开机时计为初始时刻 0），即 $t_{0s}=0$，则 $\hat{\varphi}=\frac{1}{2}(\hat{\theta}+\hat{\eta})$，连同估计的 $\hat{\beta}$ 代入 $delay_t=delay_m-\beta\cdot T_{sn}+\varphi$，即可得时延的估计值，其中 $delay_m$ 为实测到的时延值。对于这两个目标的优化，采用线性加权算法，优化目标可写为

$$\min\sum\lambda f_1+(1-\lambda)f_2$$

其中

$$f_1=delay_{rm}-\hat{\beta}T_{sn}-\hat{\xi}\cdot M_n+\hat{\theta} \quad (2)$$

$$f_2=\frac{1}{\hat{\lambda}_s}delay'_{rm}-(\hat{\lambda}_s-1)T'_d-\hat{\xi}'\cdot l-\hat{\eta} \quad (3)$$

对于线性规划方法的实现，我们采用线性优化算法编程来求目标函数 $\min\sum\lambda f_1+(1-\lambda)f_2$ 的解。

4.3 端到端时延测量高精度算法的实现

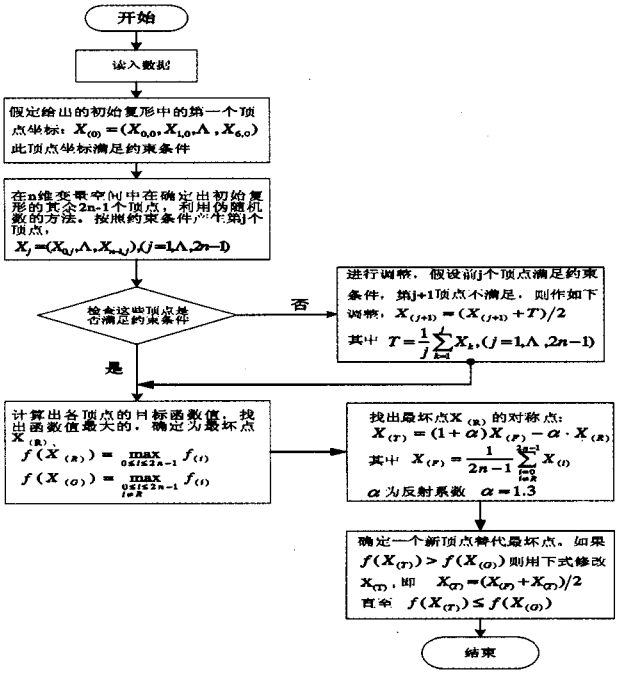

图 3 线性优化算法流程图

通过对端到端时延算法的分析，我们用 C++ 编程实现，具体算法步骤如下：

Step1：发送 IP 数据包 i 前在数据包中记录主机 A 的发送时标 $Ti,0$，同时接收方在接收到 IP 数据包 i 后则需要记录主机 B 的接收时标 $Ti,1$；

Step2：接收方主机 B 收到 IP 数据包 i 后同样也需要回复一个 IP 数据包，其中记录回复时标 $Ti,2$，原发送方主机 A 接收到该回复包记下回复包的接收时标 $Ti,3$；

Step3：求得在网络上的 3 个时延时间（未同步）：

（1）前向时延（$delay_{mi}$）：$T_{i,1}-T_{i,0}$；

（2）反向时延（$delay'_{mi}$）：$T_{i,3}-T_{i,2}$；

（3）往返时延：$T_{i,3}-T_{i,0}$；

Step4：采用线性优化算法进行时延估计，消除时钟不同步误差，获得端到端时延的真实值。算法流程图见图 3。其中，X_0、X_1、X_2、X_3、X_4、X_5 和 X_6 分别表示（2）式和（3）式中的 $\hat{\beta}$、$\hat{\xi}$、$\hat{\theta}$、$1/\hat{\lambda}_s$、$\hat{\lambda}_s-1$、$\hat{\xi}'$ 和 $\hat{\eta}$，维数 $n=7$。

5 测试实例及结果分析

我们对西电校园网和中国电信数据网、西电校园网和西安交大校园网间的两条链路进行了测试，并对测试的未同步的时延测量结果进行了时延同步。测试分组采用 UDP 协议，分组的内容为探测分组序列号和发送时刻从发送主机读取的时间值，主机的时间值可以精确到微秒级。

实例 1 源主机（西电校园网），目的主机（中国电信数据网）

测试条件：发送 1000 个 200 字节长的数据包，时间间隔为 200ms，从源主机（西电校园网）到目的主机（中国电信数据网），从 CERNET 到 Chinanet，共 15 跳。测试结果如图 4。

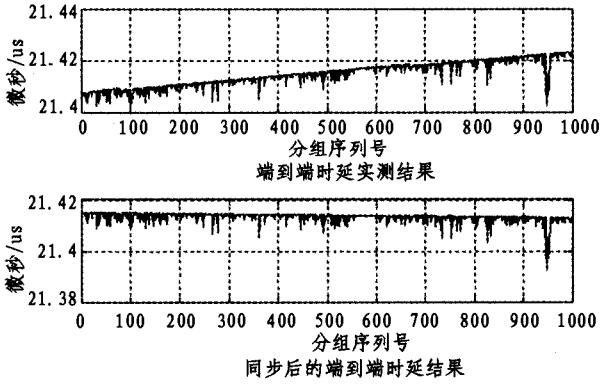

图 4 端到端时延的测试结果

其中，图 4 端到端时延实测结果中收发时钟分别采用各自的时钟，由于时钟不同步的影响，测试的结果呈现一个线性趋势。同步后的端到端时延的测试结果中采用线性优化方法估计参数，去除时钟不同步带来的影响后的时延，结果是一条近似平稳的直线。

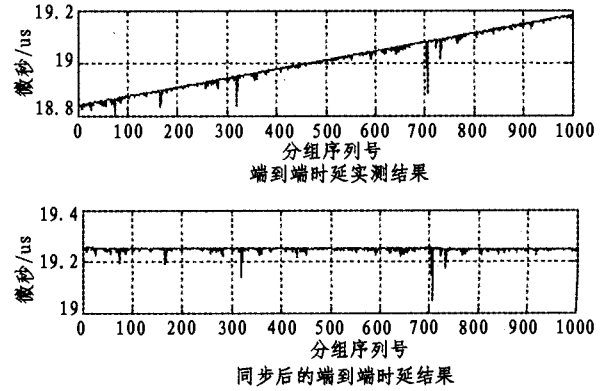

图 5 端到端时延测试结果

（下转第 128 页）

资源类型包括关系数据库、视频文件、图像文件、已存在的本体对象、Web网页(HTML,XML),以及提供信息的Web服务等。为了将这些信息资源有机地融入到本系统的一体化元数据模型中,需要依据前面介绍的元数据模型对各类信息源中包含元数据和隐含的语义本体知识进行抽取。这一过程主要采取人工注册或半自动抽取的方式,针对各种类型信息资源建立相应的wrapper,基于元数据模型抽取相关元信息和语义内容,采用XML作为元信息交互模型,生成相应的资源实例对象。

在信息系统中广泛分布的信息资源大量是采用关系数据库的方式进行存储,关系数据库采用关系模式对信息资源进行结构化存储。关系数据库是关系表的集合,一般是基于实体联系(ER)模型转换而来的,实体联系(ER)模型是对数据库概念建模的重要工具,是基于世界由一组实体的基本对象及这些对象之间的联系组成,是一种语义模型。实体联系(ER)模型与本体的概念及概念之间的关系相对应,因此我们可以对应采用本体对信息资源描述,采用扩展E-R图进行可视化描述。

本系统提供了图形化的资源元数据模型封装工具,如图4所示。

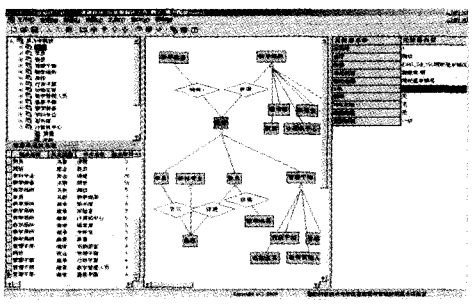

图4 元数据描述工具运行示例

结论 本文所设计的基于本体的元数据模型可以很好地支持信息资源语义描述,并通过核心元数据和扩展元数据充分兼顾广域环境下信息资源描述的通用性和特殊性的要求,并具有很好的扩展性。元数据模型定义了IS-A和Part-of的偏序关系以及语义关联关系,描述了属性以及实例与实体类的映射,能很好地支持语义相似度的计算,为基于语义相似度的语义覆盖网构建提供了基础。最后用一个原型系统对元数据模型设计的元数据规范模版加以了验证,证明是切实可行的。

参考文献

1 Maedche A. Clustering Ontology-Based Metadata in the Semantic Web

2 Wache H, Vogele T, Visser U. Ontology—Based Integration of Information—A Survey of Existing Approaches. In: Proc. of the UCAI—O1 Workshop on Ontologies and Information Sharing, Seat tie. USA, Aug. 2001. 108～118

3 陈汉华,金海,等. SemreX:一种基于语义相似度的P2P覆盖网络[J]. 软件学报,2006,17(5):1170～1181

4 刘炜. 元数据与知识本体[J]. 图书馆杂志,23(2):650～54

5 王洪伟. 基于本体的元数据模型及DAM工表示[J]. 情报学报,23(2):131～13

6 宋炜,张铭. 语义网简明教程. 北京:高等教育出版社

7 刘震,等. 面向对等网信息语义共享的元数据模型框架研究[J]. 计算机科学,2006,33(1)

8 Gruber T R. A translation approach to portable ontologies. Knowledge Acquisition, 1993, 5(2):199～220

9 Wache H, Vogele T, Visser U. Ontology-based Integration of Information —A Survey of Existing Approaches. In: Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, USA, August 2001. 108～118

10 Gruber T. A translation approach to portable ontologies. Knowledge Acquisition, 1993, 5(2):199～220

11 Borst W. Construction of Engineering Ontologies: [Phdthesis]. Enschede: University of Twente, 1997

12 Studer R, et al. Knowledge engineering: principles and methods. Data and knowledge engineering, 1998, 25:161～197

13 宋峻峰,等. OWL DL的形式化基础研究[J]. 小型微型计算机,2005,26(2):297～301

(上接第111页)

实例2 源主机(西电校园网),目的主机(西安交大校园网)

测试条件:发送1000个500字节长数据包,时间间隔为50ms,从源主机(西电校园网)到目的主机(西安交大校园网),同为CERNET网,共6跳。同样也可以看出,由于时钟不同步的影响,端到端时延实测结果呈现一个线性趋势;消除时钟不同步带来的影响后,时延变化趋于平稳。

结论 本文给出了测量网络端到端时延的方法,根据测得的结果,用线性规划的方法来消除收发时钟的初始相位差和相对频差等参数,获得端到端时延的真实值。实例测量结果表明,该方法消除了时钟不同步带来的影响,提高了测试结果的精确度,可以应用于互联网中的端到端时延测量。

参考文献

1 Van Moffaert A, De Vleeschauwer D, Janssen J. Tuning the VoIP Gateways to Transport International Voice Calls over a Best-Effort IP Backbone [A]. In: Proceedings of the 9th IFIP Conference on Performance Modelling and Evaluation of ATM&IP Networks2001 [C]. Budapest: IFIP, 2001

2 Omer G, Israel C, Moshe S. One-way delay estimation using network-wide measurements [J]. IEEE Transactions on Information Theory, 2002, 52(6):2710～2724

3 Bovy C J, Mertodimedjo H T, Hooghiemstra G, et al. Analysis of End-to-End Delay Measurements in Internet [C]. In: Proceedings of the Passive and Active Measurements Workshop(PAM 2002) [C]. Fort Collins (CO), 2002. 26～33

4 Zhu Changhua, Pei Changxing. Internet end-to-end delay dynamics [J]. Journal of Systems Engineering and Electronics, 2006, 17(3):685～691

5 Kobayashi K, Katayama T. Analysis and Evaluation of Packet Delay Variance in the Internet [J]. IEICE Trans Commun, 2004, E85-B:35～41

6 焦利,林宇,王文东,等. 一种负载均衡网络中内部链路时延推测算法[J]. 软件学报, 2005, 14(5):886～893

7 Moon S B, Skelly P, Towsley D. Estimation and Removal of Clock Skew from Network Delay Measurements [A]. In: IEEE Infocom99 [C]. New York: IEEE, 1999. 227～234